**Statement of Research Interests**

**Shicheng Guo, Ph.D.**

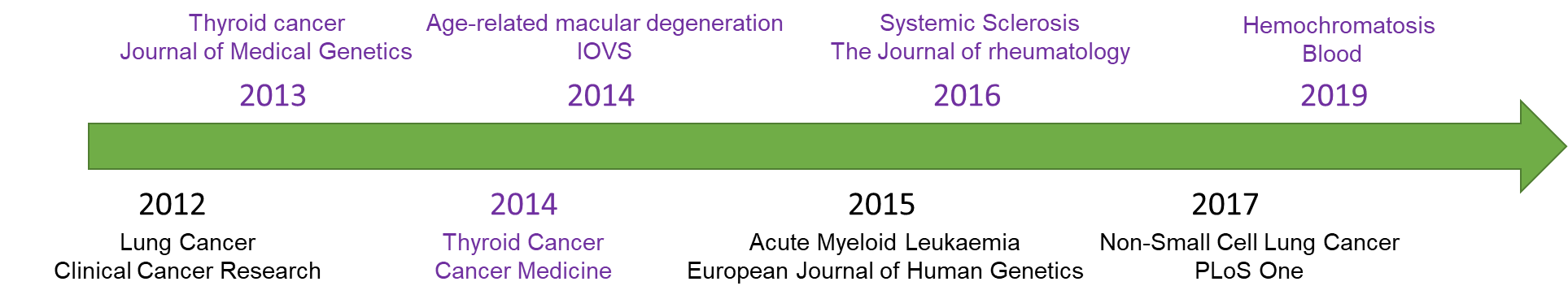
Human complex diseases may be dynamically and progressively driven by sequences of genetic-epigenetic-environmental interactions. Each of these components contributes to bring phenotypes from a susceptibility stage to clinical presentation. The most exciting ideas of precision medicine require scientists to consider numerous genetic susceptibility variants as potential baseline risk factors and track the epigenetic changes in pathogenesis of human diseases over time. Taking rheumatoid arthritis (prevalence=0.5% in US) as an example, individuals carrying HLA-DRB1\*04:01 in combination with inflammatory processes and citrullination caused by factors such as smoking, are thought to initiate an autoimmune reaction. Early molecular and epigenetic biomarkers such as rheumatoid factor (RF), anticyclic citrullinated peptide (anti-CCP) and circulating cell-free DNA methylation (cfDNAmeth) signals caused by synovial cell apoptosis preceded manifestation of symptomology. In the past 8 years, I have been working to identify genetic and epigenetic variants in multiple human complex diseases and authored >50 peer-reviewed publications, 17 of which are first/co-first/co-corresponding author publications. My extensive experience in conducting human genetics, epigenetics research combined with my bioinformatics skills in analyzing and interpreting data has prepared me to conduct comprehensive research studies in precision medicine. My future research will focus on 1) identifying and implementing detection of circulating cell-free DNA-based genetic and epigenetic biomarkers for early diagnosis, real-time monitoring and prognosis prediction; 2) Functional assessment of the genetic and epigenetic variants identified by GWAS, pheWAS, EWAS studies; and 3) developing epigenetics biomarker applying pharmaco-epigenetics (PeGx) approach to coordinate with PGx in guiding personalized treatment and precision medicine.

**Research Background**

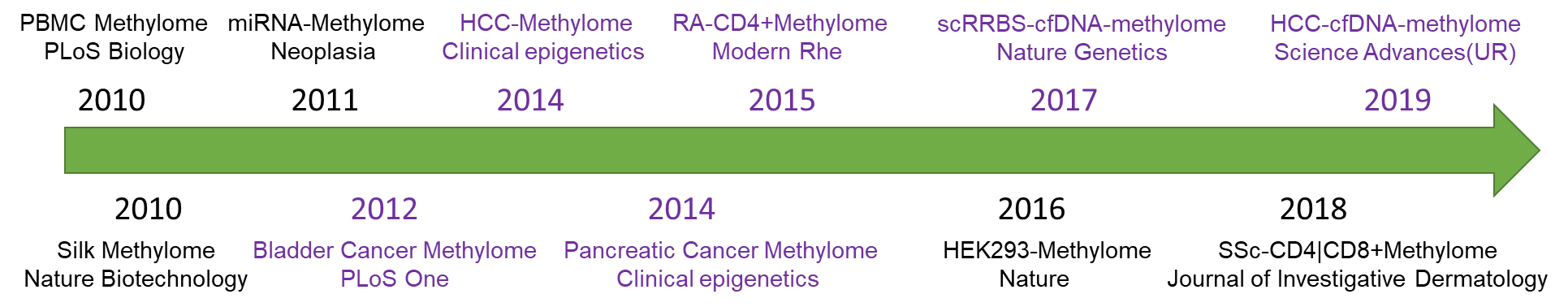
I have conducted a wide array of genetic and epigenetic studies which required me to apply both computational and basic genetic bench science. The experience from these studies has given me the ability to investigate clinically important research questions using novel and multi-faceted approaches.

**Identification of Genetic Susceptibility Genes to Human Complex Diseases.**

Early in my career, leveraging GWAS and candidate gene study, I identified multiple susceptibility genes, such as CNV within *HLA-DQA1* and *APOBEC3A/3B* for SSc and *FOXE1* for thyroid cancer. I also conducted a series of association studies investigating candidate genetic variants in miRNA for human cancer. In these study, miR-4293 was demonstrated to be significantly associated with lung cancer, and[miR-196a2/miR-499](javascript:void(0)) involved in esophageal cancer (ESCC). In addition, I implemented a novel approach termed exome-wide gene-based recessive diplotype scanning. Compared with traditional methods, our method demonstrated higher power to identify recessive compound heterozygotes. I have applied the new method to 15 diseases and analyzed genetic data available for 20,000 subjects enrolled in the Personalized Medicine Research Project (PMRP) at Marshfield Clinic. I identified a susceptibility gene, *FGF6,* for hemochromatosis. By conducting evolution analysis, protein-protein network analysis, and examining both molecular and cellular evidence, we demonstrated *FGF6* plays an important role in iron metabolism which may be involved in multiple conditions underlying diseases associated with iron metabolism. The work was accepted in Blood in 2019. The following flowchart shows the timeline for publication of outcomes of my research efforts (Purple: 1st author).



**Epigenomic Research and Epigenetic Variations in Diagnosis and Prognosis Models for Complex Diseases.** Beginning in 2009, I investigated the epigenetics of human diseases with a particular focus on DNA methylation. I participated in several large projects to build a model depicting the epigenomic architecture for human cells and tissues under normal and disease conditions. Notable work includes evaluation of the methylomes for normal human blood cells, animal model ‘silk’, CD4+ T-cells of patients with [rheumatoid arthritis](javascript:void(0)), pancreatic cancer cells, and hepatocellular carcinoma cells with W and MBD-seq. Concurrently, I identified a large number of methylation-based markers with diagnostic and prognostic implications for lung cancer, bladder cancer, and pancreatic cancer. Since DNA methylation has different patterns for different tissue types, we proposed a predictive model to map the origin of cell-free DNA fragments based on tissue-specific methylation signals. This model provides a potentially non-invasive approach for the diagnosis of solid cancers. This work was published in *Nature Genetics* in 2017. Chronology of publication of this research is shown in the following flowchart below (purple indicates 1st author):



**Current Research**

My research in Center for Precision Medicine Research consists of several studies to elucidate genetics of disease with certain non-traditional statistical methods and to identify epigenetic variant-based diagnostic and prognostic biomarkers. The details for these projects are as follows:

**A gene-based recessive diplotype exome scan to identify disease genes for 15 PMRP phenotypes.**

This project was initiated in late 2017 and is supported by Dr. Schrodi’s MCRI grant. We planned to map genes facilitating conditions including iron overload disorders, obesity, type 2 diabtetes, rheumatoid arthritis, psoriatic arthritis, multiple sclerosis, premature myocardial infraction, and obsessive compulsive disorder using a novel approach to gene-based recessive diplotype analysis based on 20,000 exome-chip data. Standard analyses applied in GWAS are well-designed to detect additive effects. However, the power for standard GWAS analyses to identify effects from recessive diplotypes is not typically high. With this approach applied to iron overload, a strong association signal was identified between the fibroblast growth factor-encoding gene, *FGF6*, and hemochromatosis in the central Wisconsin population. As shown above, this work was published in *Blood, 2019.* Currently, I am working on publication of the remaining findings in the context of type 2 diabetes (*MTNR1B* and *KIF2C*), obesity (*SPTBN5*) and rheumatoid athritis (*FSTL1*). I have completed all the bioinformatics, molecular function prediction analysis , molecular and celluar validation plans. Working together with my collabortors, I have received some exciting biological validation resulting from knock-down and up-regulation experients. Finally, I am working on the bioinformatics manuscript of ”R Packge for Compound Heterozygote Scanning to Identify Novel Disease Genes based on Exome-wide Genotype Data” which is near completion and is targeting submission to *Bioinformatics*.

**DNA Hypermethylation Mediated Epigenetic Silenced Diagnostic and Prognostic Biomarkers in Human Cancers.** This project started in early 2017 as a collaboration with Dr. Steven Schrodi and Dr. Minghua Wang. Leveraging my well-trained bioinformatics and data-mining skills, I have conducted systematic analysis integrating most well-known public data such as TCGA project, ENCODE project, Blueprint Epigenome project, Roadmap project, GTEx project (eQTL), 1000 Genome project, gnomAD dataset, and UK-biobank dataset to identify the most interesting ’differential DNA methylation region identification algorithm’ (fDMRI) for human cancers which would be considered as important and novel diagnotic and prognositc biomarkers. I applied this method and identified ~50 fDMR for esophageal cancer (ESCC) and cholangiocarcinoma. Currently, biological validation demonstrated that our method could identify novel interactions between DNA methylation and other genomic functional elements. For example, we found [epigenetic silencing of *ZNF132* mediated by methylation-sensitive Sp1 binding promotes cancer progression in ESCC](https://www.nature.com/articles/s41419-018-1236-z). Results were published in Cell Death & Disease in 2018. We will be continue to validate the remaining fDMRs and discover the mechanisms surrounding these epigenetic factors in human cancers.

**Proposed Research Program**

While human genetics opened the road to precision medicine, genetics is not enough for the achievement of precision medicine. Individual variability caused by genetics, epigenetics and environment should be concurrently considered in disease subtype, diagnosis, treatment and prevention. In addtion, our understanding to the pathogenesis of complex disease is limited since the variants identified by traditional GWAS studies can only explain very limited proportions of the overall heritability. Epigenetic variation, non-additive but complex modes of inheritance, misdiagnosis, and rare SNP and CNV effects are the most promising solutions to the missing heritability problem. In addition, recently, a serials of powerful and low-cost functional assessment and identification tools have been developed such as WES, ATAT-seq, scRRBS, BSPP, CUT&Tag, Fecal-seq, RAD-seq, cfMeDIP-seq which provide the best opportunity to investigate the role of genetic, epigenetic and interactions in the development of complex diseases. With diverse interdisciplinary training in human genetics, biostatistics, computational biology and epidemiology, my research will take advantage of my knowledge of these advanced technologies, which will be applied to the following three fields: 1) developing and implementing new approaches to identifying susceptibility, diagnosis, and prognosis-related genetic and epigenetic variants/biomarkers and novel drug targets for human complex diseases, especially cancers and autoimmune diseases; 2) applying computational and biological approaches for functional assessment of genetic and epigenetic diseasevariants identified by GWAS, EWAS and pheWAS projects; 3) developing epigenetic variants or biomarkers based pharmaco-epigenomics (PeGx) approaches that complement PGx to guide personalized medicine and treatment.

**Development and implemention of novel molecular diagnostic and prognostic approaches to human cancer.**

DNA methylation has been demonstrated to be one of the most promsing diagnostic, prognostic and pharmaco-epigenomics biomarkers for human complex disease. This may be attributable to the fact that DNA methylation is partially stable and partially dynamic, compared to genetic variation (completely stable) and mRNA (highly dynamic). DNA methylation is involved in transcriptional regulation and therefore plays critical roles in differentiation, development and disease. Given its regulatory roles, DNA methylation changes usual occur earlier than other classes of molecular variation. A large number of DNA methylation-based diagnostic and prognostic biomarkers have been identified, such as *SEPT9* and *SHOX2* which have been approved by FDA for colon cancer and early lung cancer screening. However, DNA methylation diagnostic biomarkers for other cancer types remain to be defined. In my previous publication (Nature Genetics, 2017), I have demonstrated non-invasive cell-free DNA methylation haplotype-based tissue-of-origin prediction could be applied in cancer diagnosis. With my DNA methylation expertise, I will conduct a study applying definition of novel cell-free, DNA-based, non-invasive cancer diagnosic biomarker(s) and further investigate drug response prediction. In this project, I will integrate genetic (cancer risk allele, somatic mutation), epigenetic (DNA methylation) and other informative variables, such as cell-free DNA fragment distribution, metabolics in plasma to develop a systemic prediction platform with artificial intelligence (netural network). Meanwhile, I will take advantage of the preliminary data I have collected on biobanked DNA at Marshfield Clinic Research Institute (MCRI) (or PMRP?). . Since MCRI has collected blood, plasma and serum samples since 2005, huge numbers of samples were collected before disease diagnosis, and provide a valuable resource to evaluate the earliest occurence time for detection of epigenetic/metabolic signals associated with progression of diseases such as cancer or autoimmnue diseases. In summary, I will prepare an RO1 proposal to implement this study with collaboration MCRI and UW-Madison dedicate scientists.

**Functional assessment of human genetic and epigenomic variants with computational and biological approach.**

In the past decades, GWAS, EWAS and pheWAS have identified hundreds of signficant disease-associated genetic and epigenetic loci. However, there is a huge gap between the statistically-signficant associations linking locus and human phenotypes. Functional assessment of genetic and epigenomic variants will provide the opportunity to advance understanding of the functional significance underlying disease risk and pathologenesis of complex diseases. In this study, I will propose serial functional assessment study of these human genetic and epigenomic variants with both computational and biological approaches simultaneously. In the first stage, computational assesement will conducted leveraging publically available databases or tools such as Encode, GTEx, ExAC, FANTOM, Roadmap Epigenomics, LINCS, Blueprint, RegulomeDB, BioGPS, STRING, Reactome pathway, KEGG, ANNOVA, VEP. In my previous publication (Blood, 2019), I have demonstrated that comprehensive computational and evolutionary analysis could provide highly efficient functional, protein structure-related and interactive network prediction for definition of candidate genes and variants. In the second stage, I will apply biological approaches such as CRISPR-Cas9 to validate the hypothesis generated with above method. To date, I have conducted these analysis on 15 phenotypes of 20,000 PMRP samples and identified several very interesting candidates, such as *FSTL1* in rheumatoid arthritis, *MTNR1B* in type 2 diabetes, SLX4 and SPTBN5 in obeisty. In summary, I will prepare an RO1 or KL2 proposal to continue these studies in collaboration with UW-Madison scientists.

**Phenome-wide association study of genetic variation in epigenetic factors.**

This project is a collaboration with Dr. Steven Schrodi and Dr. Mark Craven as part of the CIBM training program supported by UW-Madison and National Library of Medicine (NLM). Human complex disease is generated by the interaction between genetics, epigenetics and the environment. While the rationale for genetic association studies have been supported by different fundamental observations such as heritability estimates from twin studies, there is no fundamental research to illustrate whether epigenetic changes are involved in disease heritability, although we know that epigenetic elements are an important interface between genetics and the environment. In this study, I hypothesize that genetic variants in epigenetic genes are a proxy to infer the epigenetic involvement in phenotypes. We will apply a phenome-wide association study (PheWAS) approach to test the association between a panel of epigenetic factors against 6,221 clinical traits within the PMRP dataset. This will enable us to identify all the significant phenotypes whose pathology are potentially driven by epigenetic changes and apply the measurement of genome-wide DNA methylation levels in the corresponding phenotypes to validate the above findings. This project will also feature a collaboration with Dr. Scott Hebbring. We will share the DNA methylation dataset with Dr. Hebbring so that he will investigate the relationship between aging, telomere length and genome-wide DNA methylation. Such studies have not been previously conducted to date and the results of this work will provide insight into epigenetic architecture underlying important clinical traits. I will prepare an RO1 application to continue these studies in collaboration with UW-Madison scientists.

**Developing epigenetic variants or biomarkers based pharmaco-epigenomics (PeGx) to cooperate PGx to guide personalized medicine and treatment.**

Genetic variation in drug-metabolizing enzymes and transporters (DMET) genes in relation to drug response has been the main focus of pharmaco-genetics (PGx) laboratories. However, transcriptional regulation of DMET genes also play important roles in drug response, drug resistance and adverse drug events (ADE). In my previous studies, I have developed and implemented multiple epigenetic approach to investigate genome-wide DNA methylation and gene-panel based methylation profiling, such as single-cell reduced-representation bisulfite sequencing (scRRBS), methylation status-determined single nucleotide primer extension technique (MSD-SNuPET), and Multiplex PCR targeted bisulfite sequencing (MTBseq). Multiple studies have shown that even when we included genome-wide SNP analysis, the drug response prediction performance is still quite limited (AUC<0.8). Actually, genetic background only provides a baseline variable to predict drug response and adverse event. My hypothesis is that epigenetic status could provide emergent signals for better prediction. In this study, I plan to apply epigenetic approaches to identify blood-based DNA methylation biomarkers that supplement PGx findings to achieve a further paradigm that informs advancement of precision medicine.